Graph-Valued Regression

Presenter: Xi Chen

Han Liu, Xi Chen, John Lafferty and Larry Wasserman

Machine Learning Department

Carnegie Mellon University

Graph-Valued Regression

Multivariate Regression (supervised)

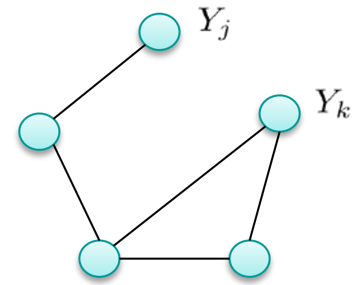
$$\mu(x) = E(Y|X=x)$$

$$Y \in \mathbb{R}^p \quad x \in \mathcal{X} = [0,1]^d$$

Undirected Graphical Model (unsupervised)

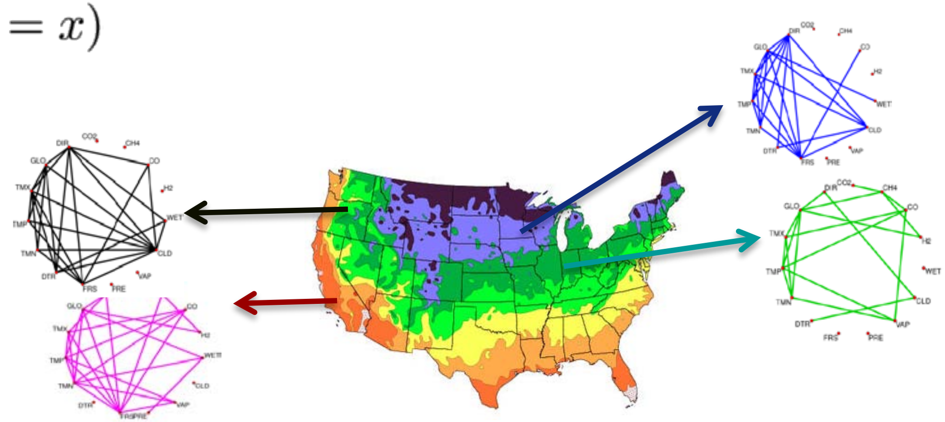
$$\text{Graph}(Y) = (V, E)$$

$$(j,k) \notin E \Leftrightarrow Y_j \perp\!\!\!\perp Y_k \,|\, \text{rest}$$



Graph-Valued Regression: $\text{Graph}(Y|X=x)$

Application: complex data analysis



Learn Graph(*Y*) via Graphical Lasso

$$Y \sim N_p(\mu, \Sigma) \qquad \Omega = \Sigma^{-1}$$

$$\Omega_{jk} = 0 \Leftrightarrow Y_j \perp\!\!\!\perp Y_k \,|\, \text{rest} \Leftrightarrow (j,k) \notin E$$

[Yuan & Lin 07]

$$\widehat{\Omega} = \arg\min_{\Omega \succ 0} \{ \underbrace{\text{tr}(S\Omega) - \log|\Omega|}_{-\text{ Log-likelihood}} + \underbrace{\lambda \|\Omega\|_1}_{\text{Sparsity}} \}$$

– Log-likelihood   Sparsity

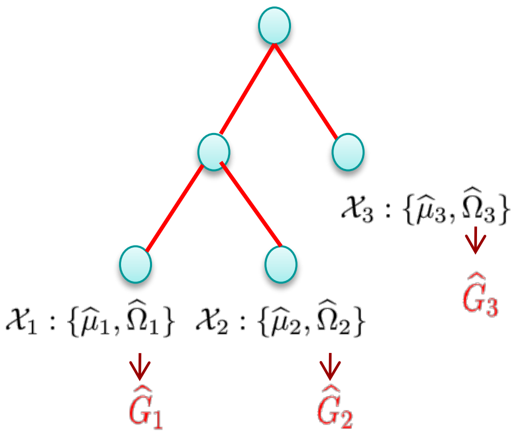
S: Sample Covariance Matrix

Go-CART (Graph Optimized CART)

Co-CART (Graph-Optimized CART):

[Scott 2004]

$$Y \mid X = x \sim N_p(\mu(x), \Sigma(x))$$



Partition based estimator:

Input data: $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_i \in \mathcal{X} = [0,1]^d$ and $y_i \in \mathbb{R}^p$

1. **Dyadic Decision Tree based Partition:** $\mathcal{X} = \bigcup_{j=1}^m \mathcal{X}_j$
2. **Estimate Graphs:** $\widehat{G}_j$ via *Graphical Lasso* based on $\{y_i : x_i \in \mathcal{X}_j\}$

$$G(Y|X = x) = \widehat{G}_j, \quad \forall x \in \mathcal{X}_j$$

Co-CART Estimators:

1. Penalized Empirical Risk Minimization Estimator: Minimize negative conditional log-likelihood + tree-induced penalty
2. Held-out Risk Minimization Estimator: Minimize held-out negative conditional log-likelihood

Statistical Property:

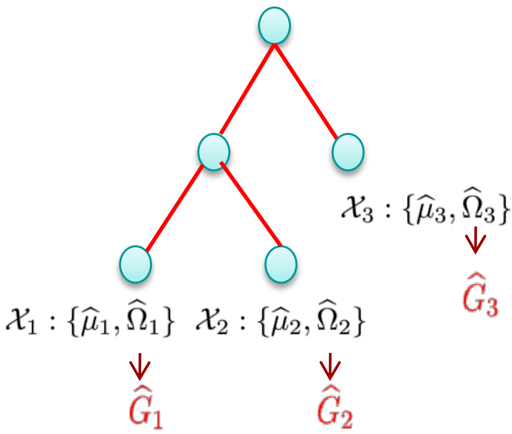
- Oracle inequality
- Tree partition consistency : $\mathbb{P}\left(\Pi(T^*) \subset \Pi(\widehat{T})\right) \to 1$

Go-CART (Graph Optimized CART)

Co-CART (Graph-Optimized CART):

[Scott 2004]

$$Y \mid X = x \sim N_p(\mu(x), \Sigma(x))$$



Partition based estimator:

Input data: $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_i \in \mathcal{X} = [0,1]^d$ and $y_i \in \mathbb{R}^p$

1. **Dyadic Decision Tree based Partition:** $\mathcal{X} = \bigcup_{j=1}^{m} \mathcal{X}_j$
2. **Estimate Graphs:** $\widehat{G}_j$ via *Graphical Lasso* based on $\{y_i : x_i \in \mathcal{X}_j\}$

$$G(Y|X = x) = \widehat{G}_j, \quad \forall x \in \mathcal{X}_j$$

Co-CART Estimators:

1. Penalized Empirical Risk Minimization Estimator: Minimize negative conditional log-likelihood + tree-induced penalty
2. Held-out Risk Minimization Estimator: Minimize held-out negative conditional log-likelihood

Statistical Property:

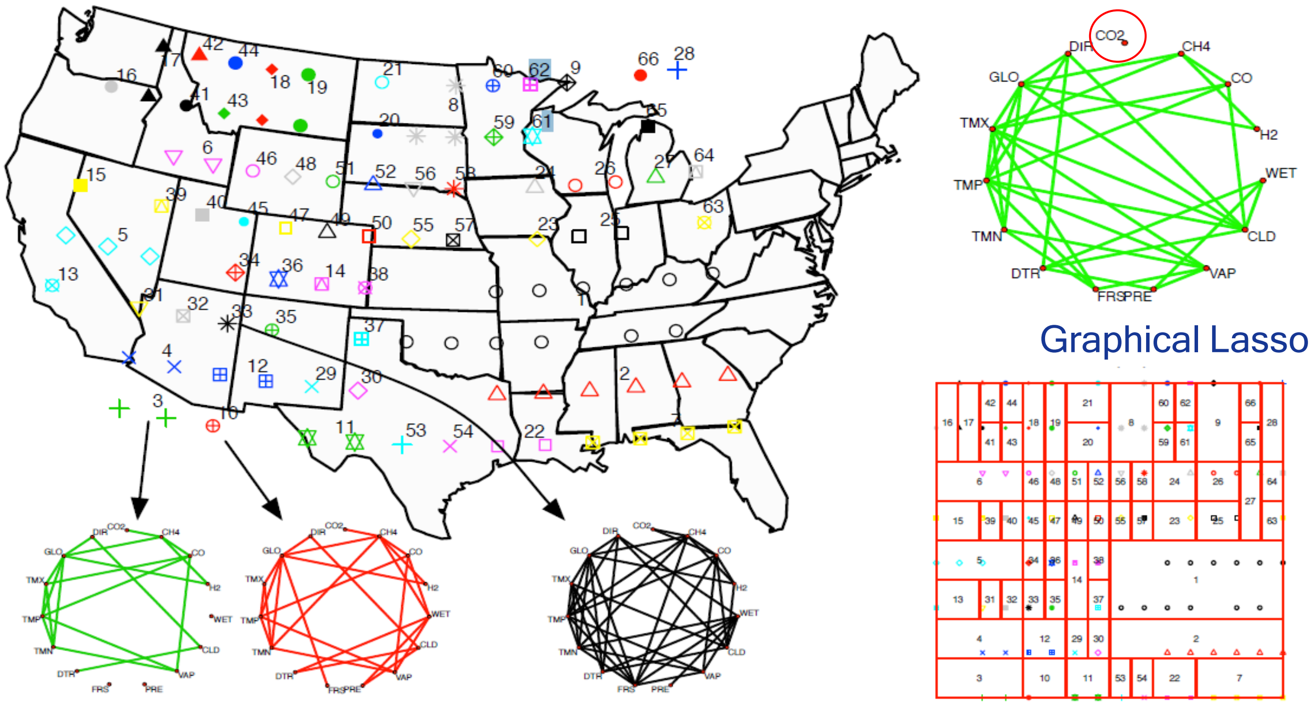
- Oracle inequality
- Tree partition consistency : $\mathbb{P}\left(\Pi(T^*) \subset \Pi(\widehat{T})\right) \to 1$

Experimental Results

Real Climate Data Analysis

held-out risk minimization with a greedy tree learning algorithm

100 locations of U.S.; Monthly observation 1990 ~ 2002 (training: 1996~2002) ; 15 factors (greenhouse gases, e.g. CO_2; temperature, climate, solar radiation)



Observations:

(1): Graphical lasso, no edge connect CO_2 with solar radiation factors; Co-CART, there is.

(2): Graphs along the coasts are more sparse than the ones in the mainland.